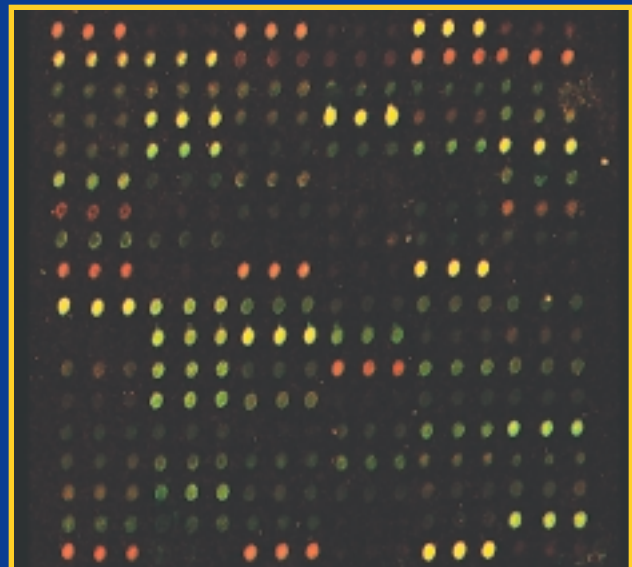




National Plant Genome Initiative

Progress Report
October 1999

National Science and Technology Council
Committee on Science
Interagency Working Group on Plant Genomes



About the National Science and Technology Council

President Clinton established the National Science and Technology Council (NSTC) by Executive Order on November 23, 1993. This cabinet-level council is the principal means for the President to coordinate science, space, and technology policies across the Federal Government. NSTC acts as a “virtual” agency for science and technology to coordinate the diverse parts of the Federal research and development enterprise.

The NSTC is chaired by the President. Membership consists of the Vice President, Assistant to the President for Science and Technology, Cabinet Secretaries and Agency Heads with significant science and technology responsibilities, and other senior White House officials.

An important objective of the NSTC is the establishment of clear national goals for Federal science and technology investments in areas ranging from information technologies and health research, to improving transportation systems and strengthening fundamental research. This Council prepares research and development strategies that are coordinated across Federal agencies to form an investment package that is aimed at accomplishing multiple national goals.

To obtain additional information regarding the NSTC, contact the NSTC Executive Secretariat at 202-456-6100.

Note: This document does not represent the final determination in an overall Administration budget decision-making process. The programs presented in this report will have to compete for resources against many other high-priority Federal programs. If these programs compete successfully, they will be reflected in future Administration budgets.

Cover Photo: A microarray image of plant gene expression studies, courtesy of Dr. George Karlin-Neumann, Stanford University. Microarray experiments like this one allow a global analysis of gene expression in plants responding to different stimuli such as temperature, drought or pathogen attack. The information gained from this type of study will be essential for researchers to zero in on the groups of genes that work together to modulate particular plant responses. Many of these responses are of agricultural importance. Microarray analysis is thus going to be an important part of the design of tomorrow's crops.

THE WHITE HOUSE

WASHINGTON

December 2, 1999

Dear Colleague:

This report provides an update from the National Science and Technology Council (NSTC) Interagency Working Group (IWG) on Plant Genomes on the progress made by the scientific community toward implementing the goals of the NSTC's National Plant Genome Initiative (NPGI). The rate of progress has been impressive. Since January 1998 when the NSTC launched the Initiative, the target date to complete the sequence of the *Arabidopsis* genome has been accelerated from 2002 to late 2000. The United States is now a participant in the International Rice Genome project and tens of thousands of expressed sequence tags for numerous plant species have been deposited in the publicly accessible database GenBank.

As exciting as this progress has been, much remains to be done. We are now beginning to understand plant genomes and how they might be better harnessed to develop new crop varieties with improved agronomic and nutritional qualities, with the ability to produce valuable biomaterials and chemicals, and with benefit to the environment.

For the NPGI to be successful, the IWG will need to continue to promote the principles contained in its 1998 report: partnerships with industry and the international community, coordination among agencies, scientific merit and competitive review as the basis for funding decisions, open access to research resources, and stakeholder input when setting priorities and research goals. In keeping to these principles, the Initiative will continue to serve the public's interest by advancing cutting-edge plant biology.

Sincerely,

A handwritten signature in black ink that reads "Neal Lane". The signature is written in a cursive, slightly slanted style.

Neal Lane
Assistant to the President
for Science and Technology

**Interagency Working Group on Plant Genomes
Committee on Science
National Science and Technology Council**

Co-Chairs

Mary E. Clutter
Assistant Director for Biological Sciences
National Science Foundation

Eileen Kennedy
Deputy Under Secretary
Research, Education, and Economics
U.S. Department of Agriculture

Members

Elke Jordan
Associate Director
National Human Genome Research Institute
National Institutes of Health

Judy St. John
Associate Deputy Administrator
Agricultural Research Service
U.S. Department of Agriculture

Sally Rockey
Deputy Administrator
Cooperative State, Research, Education
and Extension Service
U.S. Department of Agriculture

Gregory L. Dilworth
Division Director
Office of Basic Energy Sciences
U.S. Department of Energy

Clifford Gabriel
Deputy to the Associate Director
Office of Science and Technology Policy

Noah Engelberg
Program Examiner
Office of Management and Budget

Acknowledgement

The Interagency Working Group on Plant Genomes acknowledges the assistance of Sharlene Weatherwax (Office of Basic Energy Sciences), Jane Silverthorne (National Science Foundation), Ed Kaleikau (Cooperative State Research, Education and Extension Service), and Machi Dilworth (National Science Foundation) in preparation of this progress report.

**National Plant Genome Initiative
Progress Report
October 1999**

National Plant Genome Initiative - to understand the structure and function of genes in plants important to agriculture, environmental management, energy, and health

“Genomics” is the study of a genome, which refers to the complete genetic makeup of an organism. The conceptual revolution sparked by genomics and related sciences is dramatically changing the field of plant biology. Recognizing the enormous scientific and economic potential for the future of US biotechnological advances, a National Plant Genome Initiative (NPGI) program was initiated in 1997 by the Office of Science and Technology Policy (OSTP) through its National Science and Technology Council (NSTC), at the request of Congress. The OSTP established an interagency working group (IWG) on plant genome research, comprised of representatives from NSF, USDA, DOE, NIH, OSTP, and OMB. The IWG published a five-year plan and rationale for the NPGI in January 1998. The long-term goal of the initiative is “to understand the structure and function of genes in plants important to agriculture, environmental management, energy, and health”.

The Initiative - Year One:

In the January 1998 publication, the IWG identified six goals that focused on building the requisite plant genome research infrastructure. Specific direction was given that “the NPGI should be viewed as a long-term project, governed by a plan that will be updated periodically, based on assessment of success in reaching critical milestones and of the rapidly changing state of the art.” The initial plan is being implemented by the participating agencies to the extent possible under current funding levels. Progress made since January 1998 toward each of the six goals is summarized below, along with future plans to further the goals.

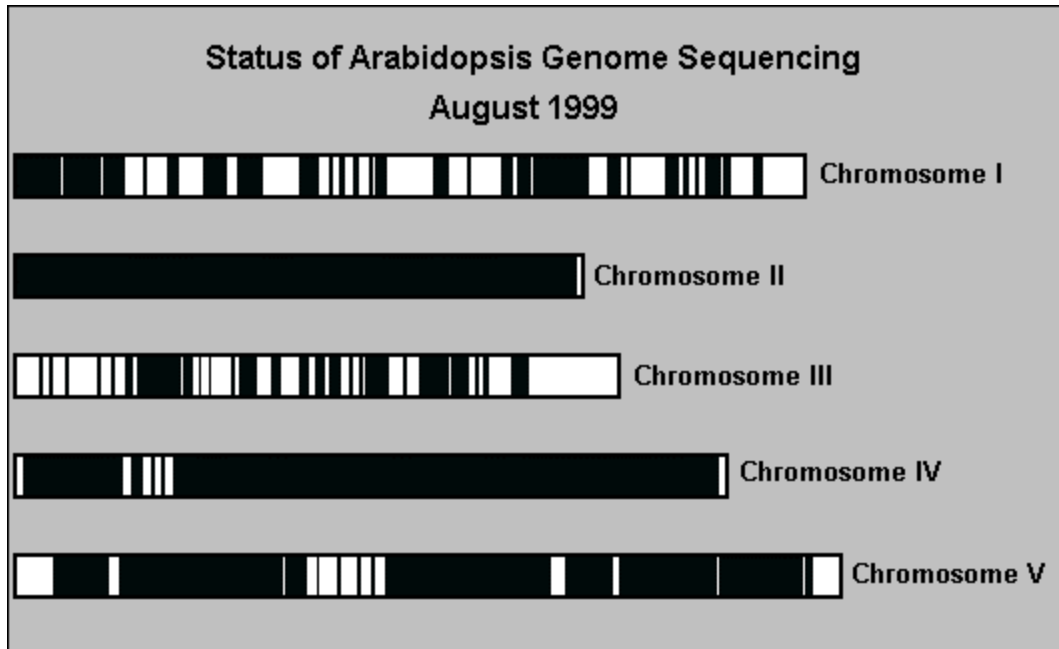
Goal 1. Sequencing the *Arabidopsis thaliana* and Rice Genomes – “Accelerate complete sequencing of the *Arabidopsis* genome and participate in the international effort to sequence the entire Rice genome.”

A. Sequencing the *Arabidopsis thaliana* Genome

Progress to Date

At the time of the establishment of the NPGI, the US component of effort to sequence the entire genome of *Arabidopsis* was well underway, having been started in 1996 as a joint program by the NSF/USDA/DOE. The US effort was part of an international consortium to sequence the *Arabidopsis* genome by the year 2004. The IWG predicted in its January 1998 report that “It is anticipated that the genome of *Arabidopsis* could be completed in the year 2000 with sufficient funding”.

To implement this NPGI goal, in FY1998 the NSF/DOE/USDA held a competition to accelerate completion of *Arabidopsis* genome sequencing. Three US groups received awards. The international partners also received additional funding, and the consortium began increasing its output in late 1998. As of August 1999, nearly 70% of the genome has been sequenced and the data released in GenBank. It is now expected that the sequencing of the *Arabidopsis* genome will be complete by the end of 2000.



[From AtDB at Stanford University. Solid regions indicate finished sequences in GenBank.]

Future Plans

When sequencing is completed at the end of the year 2000, *Arabidopsis* will be the first flowering plant genome to be completely sequenced. We will know the sequence of approximately 25,000 genes that make up a basic set of genes for a fully functional flowering plant. However, we will not know the function of the genes. For the sequence data to be fully useful to plant genome researchers and the plant science research community, the sequence data must be further refined through the process called annotation. The initial annotation accompanying the genome sequence data simply identifies genes along the entire genome. In addition, approximately 50% of the gene sequences in the database will contain a second level of annotation where their potential functions are postulated based on similarities to other genes as determined by the use of software and other computational methods. This second phase annotation will provide a hypothesis that must be verified subsequently by experimental means, leading to a comprehensive third level of annotation.

The next logical step for the plant genome research community is to complete the second phase annotation for 100% of the gene sequences, and to add the third level

annotation, namely to assign confirmed functions for all of the genes in the *Arabidopsis* genome. This represents a major effort that will require development of new software tools and other high throughput techniques that enable rapid processing of large amounts of data and information. But the community of plant biologists predicts that with a coordinated systems approach and adequate funding, the goal can be accomplished in 10 years. This effort will identify groups of genes involved in a specific process (e.g., all the genes involved in response to a fungal pathogen attack), or indicate a type of function for a specific gene (e.g., a gene involved in transporting ions across membranes).

Results from this effort will provide a solid foundation and a springboard for plant biologists to conduct functional genomics research by which to relate the function of individual genes to how plants grow, develop, and perform various life processes.

B. Sequencing the Rice Genome

Progress to Date

In order to implement this goal, the IWG has developed an interagency program to sequence the rice genome, which is integrated into an international effort led by the Japanese Rice Genome Project. The ultimate goal of the rice project is to sequence the entire rice genome. Rice belongs to the family of grasses, which are one of the most diverse groups of plants. Grasses include the world's major food crops such as barley, corn, sorghum, sugarcane and wheat. Rice has the smallest known genome of all grasses, with 430 million base pairs of DNA divided into 12 chromosomes. Since most grasses have common sets of genes, what is learned from the study of the rice genome will be immediately applicable to the other grasses.

An interagency program solicitation released in January, 1999 by USDA, NSF and DOE, resulted in an agreement to jointly fund two U.S. projects totaling \$12.3M over 3 years. The U.S. efforts will be coordinated with the international effort, whose goals are set by the International Rice Genome Sequencing Working Group. Currently the members include scientists representing Canada, China, EU, France, India, Japan, Korea, Singapore, Taiwan, Thailand, and the US. The working group is responsible for planning the most efficient means of completing the rice sequencing project to avoid duplication of efforts and maximize overall progress. It has its own public Web site (<http://www.staff.or.jp/Seqcollab.html>). US participation will ensure that the international rice genome sequencing project will follow standard policy for public genome sequencing projects on rapid data release and free information sharing, and that the international rice effort will have access to constantly evolving technologies and strategies in high throughput genome sequencing and data management.

Future Plans

It is anticipated that the funding of this aspect of the initiative will continue until the rice sequence is complete. A time-table published by the international working group in February 1999 indicates that the completion is expected by 2008 based on the current technology, and that over 1/3 of the genome will be sequenced in three years based on

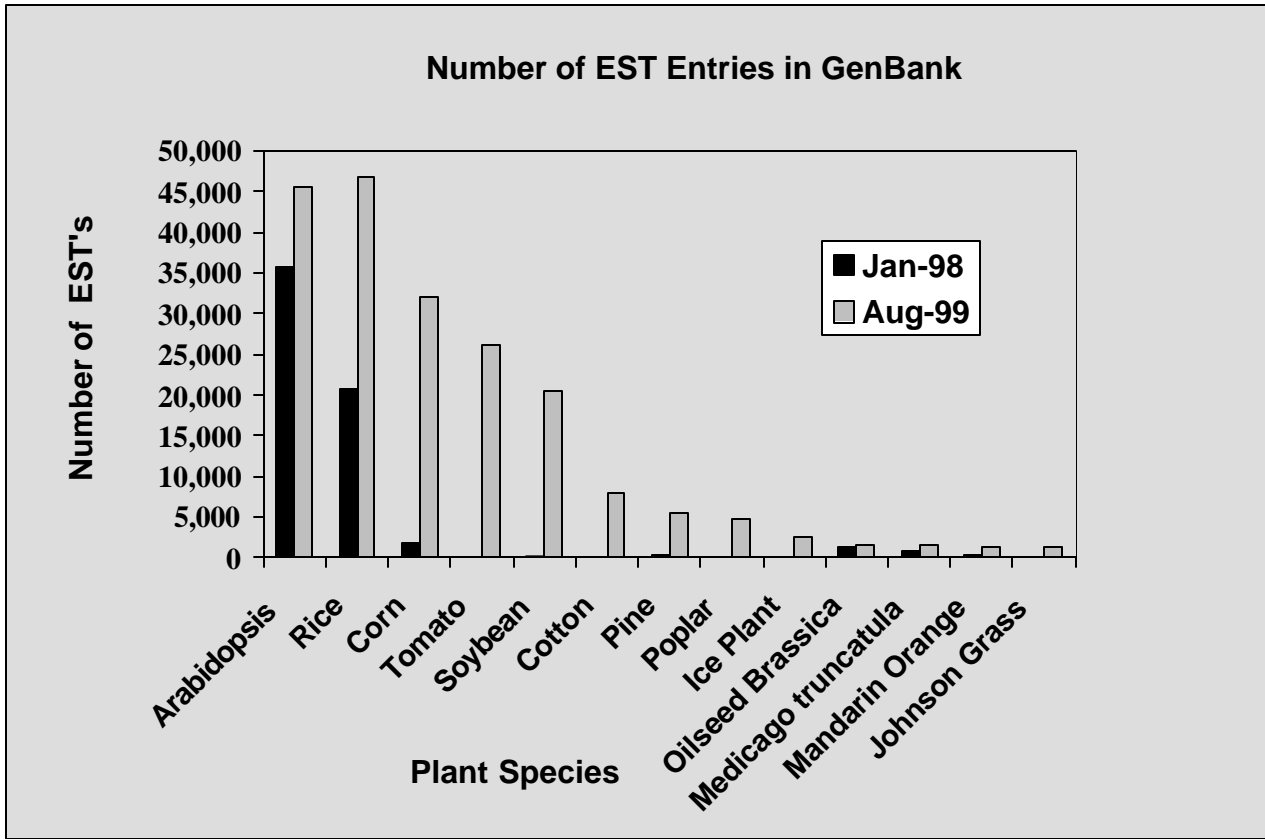
the current funding commitment from various national and international programs. They expect that predicted rapid advances in sequencing technologies will most likely allow completion of the rice genome sequence significantly earlier than this prediction. This prediction is consistent with the statement in the IWG January 1998 report, that “It is anticipated that the genome of rice could be completed in the year 2004, with sufficient funding.”

Goal 2. Structural Genomics – “Elucidating the structure and organization of genomes with the initial focus on developing physical maps and construction of expressed sequences (EST’s) for 10-12 crop species as well as a few “exotic” species.”

Progress to Date

During the past 18 months, many new projects have been supported that aim at developing the biological research resources that are essential for elucidating the structure and organization of complex plant genomes. Collaborative structural genomics research projects are now underway for many widely grown plants, including barley, canola, corn, cotton, lettuce, loblolly pine, peach, potato, poplar, rice, sorghum, soybean, sunflower, tomato, and wheat. These projects have provided the research community with genetic maps, physical maps, EST’s, DNA clone libraries, and mutant populations with specific genes tagged. The progress is most evident in the number of EST sequences deposited in GenBank, a public database maintained at the National Library of Medicine. In January of 1998, only *Arabidopsis*, corn and rice had more than 1,000 entries, with above 20,000 entries for *Arabidopsis* and rice. In August of 1999, there were 13 plant species with over 1,000 EST’s entered into the dbEST database in GenBank, with corn, tomato and soybean EST entries numbering above 10,000. As for “exotic” species, two of the projects supported by the NPGI include the construction of a small number of EST’s as an integral part of their respective experimental plans to identify drought tolerant genes in iceplant or genes involved in legume-nitrogen fixing microbe interactions in *Medicago truncatula* (a non-commercial legume).

The availability of these various resources have changed the way individual laboratories conduct their research, by allowing them to pursue biology-based research efficiently in a cost-effective manner.



[Data from dbEST/GenBank: Jan 98 data courtesy of Dr. E. Retzel, University of Minnesota.]

Future Plans

Support of this aspect of the initiative should be augmented to include more activities of unique interest to the agricultural and bioenergy research communities, as well as a few more representative “exotic” plant species in order to search for useful genes not present or expressed in economically important plants. In addition, syntenic maps will continue to be refined for a group of plant species, which gives us useful information about the organization of plant genomes.

During the past 18 months, the cost of sequencing has decreased and more and more institutions have established genome sequencing facilities or genome research centers. As a consequence, the number of institutions capable of conducting structural genomics studies is increasing. Also, the proliferation of EST’s and physical mapping information has made it possible and more efficient for scientists to integrate structural genomics studies with functional genomics studies. It is anticipated that the focus of the NPGI research area will shift toward functional genomics in the near future.

Goal 3. Functional Genomics –“ Involves identification of functions for gene sequences, including determining expression patterns for pathways or networks of genes under specific environmental conditions or during specific developmental stages.”

Progress to Date

This goal addresses directly one of the major reasons for supporting the NPGI. Infrastructure building activities such as sequencing the genomes of *Arabidopsis* and rice, physical and genetic mapping of various plant genomes, or developing software to effectively use the massive amounts of genome data being generated, are all essential in providing materials and tools to increase our understanding of the molecular basis of genes involved in important plant processes. The NPGI targets genes obviously important to plant production and productivity, such as those coding for disease and stress resistance, seed development, grain-quality traits, carbon allocation, flowering time, biomass production, and synthesis of valuable fuels and chemicals. Also included are those genes that regulate other genes, which are difficult to identify by traditional experimental approaches.

Research in this critical area is still in its inception, but a number of new activities have been initiated that will result in new discoveries and increased understanding of plant gene structure and function. Topics of some newly initiated research projects include:

- Identification of the key genes involved in drought and salt stress tolerance in plants.
- Identification and characterization of the network of genes involved in cotton fiber development.
- Quality of oil and starch in corn seeds.
- Genes involved in wood formation in trees.
- Genomic basis of plant-microbe interactions in biological nitrogen fixation.
- Identification of genes involved in disease resistances in major crops.
- Mechanisms controlling the micronutrient content of plants.
- Genomic basis for various developmental processes including development of fruits, tubers, roots, and flowers.

Future Plans

It is anticipated that funding of this aspect of specific projects already initiated will continue. It is further anticipated that the focus of the NPGI will shift toward functional genomics. As fast as the results from the structural genomics and infrastructure building efforts become widely available, they are being utilized by plant biologists in all subdisciplines, from biochemists to agronomists and from plant physiologists to ecosystems experts. This trend is expected to increase through the foreseeable future. Although specific functions that would be studied will depend on the research proposals received by the IWG agencies, each agency will likely focus on the functions that are appropriate for its mission.

Goal 4. Technology Development – “For technologies and methods specifically designed to advance plant genomics.”

Progress to Date

The plant genomics community has been quick to adapt and utilize the sizable Federal investment in genomic technologies by the Human Genome Project. At the same time, there are unique opportunities and needs for the plant genome research community. One of the areas the IWG identified as a promising technology in the January 1998 report is DNA chips/microarrays. These technologies are potentially powerful analytical tools to study the total expression patterns of genes under specific conditions. Many NPGI projects including corn, cotton, soybean, tomato, potato, sorghum and wheat genome projects include plans to use this technology. In addition, support is being provided for projects that are designed to develop a new generation of microarray/chip technologies, to develop software tools to analyze the expression patterns obtained, and to create information technologies such as search engines needed for the scientific community to access and utilize data resulting from the expression studies.

Other technologies that are being developed include new mapping methods, imaging systems that will allow investigators to observe cellular or molecular function of genes in real time, novel cloning vectors, methods to tag genes of interest in the whole plant, and reverse genetics technologies where one can determine the function of a gene from its DNA sequence.

Future Plans

The NPGI will continue to encourage the community to develop new technologies and methods that will push the frontier of plant genomics further. Advances in the field of genomics have been intertwined with advances in technologies including automated data generation and analysis that have allowed high throughput biology, miniaturization of analytical instrumentations that has increased cost-effectiveness, and the entire sub-field of bioinformatics that has provided tools to monitor, analyze, access and utilize all types and massive amounts of genomic research data. The IWG expects that the NPGI will continue to work with the other genome projects and contribute to the advances in genomics in general and in plant genomics in particular.

Goal 5. Distribution and Use of Genome Data and Resources – “Extensive data and resources generated by the NPGI must be shared and utilized.”

Progress to Date

All IWG agencies require that information and materials resulting from their support must be made available in a timely and easily accessible manner. All sequence data from the NPGI are being deposited rapidly in GenBank, the international repository for sequence data, and in turn being made widely available to the scientific research community. There are also organism-specific databases for most of the major species of plants. Many of them have been supported by the Agricultural Research Service, while some are being created or expanded under the NPGI. DNA clones, seeds, and

populations of mutant plants are deposited in public stock centers such as the *Arabidopsis* Biological Resource Center and the Maize Cooperative Stock Center or distributed via other vendors such as the American Type Culture Collection at a reasonable cost. All stock centers and distribution centers provide extensive user support. All large plant genome research centers have a public web-site where research results and information are shared with the general research community.

Future Plans

A great challenge is posed by the immense volume of information being generated worldwide not only from plant genome projects, but also from plant research in general. How can these data be rendered easily accessible and usable to an ever-increasing and broadening community of users, ranging from those in plant and general life science research, to policy makers to educators and their students? With the rapid proliferation of plant data collections, the traditional centralized approach of collection and distribution of all plant data is no longer practical or even desirable. New approaches to the coordination of the many disparate and massive datasets that will allow cross-collection access in a seamless manner will be sought through the NPGI.

The NPGI will also support community-driven development of standardized nomenclature, development of minimum common principles of database design, and the development of specific software tools designed to facilitate query across multiple databases. This will require innovation at the community level as well as at the technical level. In particular, the community level innovation will require real cooperation among the scientists generating data and coordination among each of the funding agencies supporting the generation of plant data.

Goal 6. Outreach and Training – “In order to ensure rapid transfer of genomic information and technologies to their end users, outreach activities should be an integral part of the overall plan for the NPGI.”

Progress to Date

Plant genome research provides a unique training opportunity for students at all levels. As a field at the cutting edge of biology, it provides an opportunity for young students to be exposed to the forefront of science as well as new paradigms in biological research. Because of its interdisciplinary nature, plant genome research is an ideal activity for a range of researchers, including biologists, computer scientists, engineers, chemists and others, to work together in a collaborative environment.

Many of the new projects funded since the inception of the NPGI, including all the NSF funded projects, involve training of undergraduate students, graduate students and postdoctoral fellows in some aspects of plant genome research. Undergraduate students are especially suited to be trained in the process-oriented aspects of genomic research such as genome sequencing or EST projects. These projects will expose students to a broad range of basic experimental protocols: extracting DNA's, constructing libraries, subcloning pieces of the genome, running the DNA sequencers, using various software to interpret raw data, depositing the data into the public

databases, and retrieving and using the information in the databases. Graduate students and postdoctoral fellows receive more specialized training where they acquire the skills to integrate information technologies into their biology research. Also, graduate students and postdocs learn to interact with their colleagues located outside of their immediate institutions and/or their fields of specialization. In addition, some of the IWG agencies support plant genome research training through existing/continuing base activities such as the Presidential Early Career Awards for Scientists and Engineers (PECASE) program, the postdoctoral research fellowships programs, and various workshops/summer courses.

The IWG believes that the NPGI should be able to make a significant contribution to raise the public's awareness of new scientific developments resulting from plant genome research, by providing timely and accurate information that is based on solid scientific evidence. A step toward this goal has been taken by some of the new NPGI awards. Examples of outreach activities include: providing high school teachers hands-on research experience in plant genome research; participating in local outreach programs where participating scientists visit local class rooms or civic groups to talk about plant genome research; holding workshops for agricultural extension agents to inform them about plant genome research.

Future Plans

It is anticipated that the IWG and its members will continue to encourage, participate in and actively support education and training activities. The NPGI investigators will become increasingly involved in public dialogue about the broad societal impact of plant genome research through participation at public forums and conferences involving the end users of the NPGI research results.

Other Issues:

International Partnerships - As previously mentioned, the *Arabidopsis* and rice genome sequencing projects are multinational coordinated projects, whose participants are supported by their own national programs and guided by representatives of the scientific community. These projects share information and exchange ideas freely among the participants as well as with the rest of the scientific community.

The NPGI promotes and encourages international collaborations. Some of the newly funded plant genome projects such as the wheat genome and the *Medicago truncatula* genome projects have international counterparts in Europe. The potato project works closely with both its European counterpart and the international potato center in Peru. These collaborations benefit all by expanding the scientific horizons beyond institutional, disciplinary, geographical and cultural boundaries. In addition, the international partnerships provide opportunities for US researchers and students to obtain foreign research experience, which is important in any increasingly global field of science such as genomics.

Industrial Partnerships - Various private sector concerns have reacted differently to the NPGI. Some growers associations such as the American Soybean Growers Association, the Sugarcane Association and Cotton Incorporated have contributed funding for the publicly funded genome projects that benefit them directly. Large agricultural companies are mostly providing modest levels of funding or in-kind support for specific projects on an individual basis. At least one company, Novartis, has participated directly in two corn genome projects and the rice genome research project.

One model for effective industrial partnerships might be the recently formed, nonprofit SNP Consortium, provided with \$46M by ten international pharmaceutical companies and the Wellcome Trust philanthropy of the U.K. The consortium will support a collaboration between leading U.S. and U.K. academic research centers to create a public database of defined genetic markers. These SNPs (single nucleotide polymorphisms) can serve as landmarks along the map of the human genome and can be used as analytical tools to identify variability among the human genetic code. The data and the SNP map will be shared freely with the public without any restrictions to the users. A similar alliance with the goal of providing publicly accessible, fundamental datasets on plant genomes would help advance the field of plant genomics overall.

Broader Impacts - In addition to building the scientific foundation for the future of plant sciences and plant-based industries, the NPGI takes into consideration its broader impacts to general scientific infrastructure of the nation.

Intellectual property rights: The January 1998 IWG report discussed the issue of intellectual property rights (IPR). IPR issues relevant to the NPGI relate to sharing of the information and materials resulting from the NPGI awards. Usual Federal technology transfer policies and institutional IPR policies are being followed by the NPGI awardees. Therefore, the awardee institutions retain the rights to the results of the Federally funded research, but those results must be shared with the public in a timely manner at a reasonable cost. This leaves plenty of room for individualized modes of implementation. Indeed, information sharing and material transfer policies of the NPGI awardees vary, including those entirely free with open and immediate release of all data and materials, and those requiring the use of material transfer agreements (MTAs).

Access to research resources resulting from research conducted at academic institutions is an issue of great concern to the entire scientific community. The National Science and Technology Council's Subcommittee on Biotechnology has commissioned the National Research Council to address this issue in depth. The IWG is working closely with the NSTC Subcommittee to coordinate the NPGI's policies on data release and material sharing.

Broader participation in NPGI by the scientific community: The ultimate success of the NPGI will be judged by how well new technologies and knowledge are utilized by the rest of the scientific community to advance all fields of plant sciences from basic research to applied sciences and commercial developments. For the past 18 months, new NPGI awards have focused primarily on building fundamental tools for plant genome research, by a large group of investigators

(or a virtual center). As the basic sets of tools become available, it should be possible for all academic institutions large and small to participate in the NPGI. The NPGI also creates enormous opportunities for scientists with specific missions, be it in plant breeding or bio-based products. Existing programs at the IWG agencies are well equipped to manage proposals from scientists across the US who have innovative ideas to advance the field of plant sciences using the information being generated by the NPGI.

Updates on Funding

The January 1998 IWG report estimated that a minimum new investment of \$320M, and more realistically \$400M, for five years (FY98-02) would be needed to meet the goals of the NPGI. Support provided to the NPGI through the IWG agency programs for FY98 – 00 will fall short of this goal.

Nevertheless, the investment that has been made has generated and will continue to contribute significant amounts of new discoveries, information, tools and materials as summarized above in this report. These results open up opportunities to fundamental plant biology researchers as well as researchers who are interested in translating them into practical applications matching the central missions of the IWG agencies.

Considering that the strength of the US research enterprise is based in large part on the multiplicity of funding sources, the IWG recommends that additional investments be made at all the IWG agencies to capitalize on the momentum that currently exists.

Recommended Investment for the Next 3 Years (FY2000-2002):

- \$50M for continued participation in sequencing of rice genome and complete annotation of *Arabidopsis* genome sequences.
- \$20M for EST's for additional plant species both economically important and related species.
- \$80M for cDNA and genome sequencing of specific regions of chromosomes of selected plant species.
- \$100M for informatics and data managements.
- \$230M for functional genomics research.
- \$20M for education, training and outreach activities.

Glossary

Base Pair (bp): Nucleotide bases (adenine, thymine, guanine and cytosine) are the building blocks of DNA. Two molecules of nucleotide bases held together by weak bonds. Two strands of DNA are held together in the shape of a double helix by the bonds between base pairs. The number of base pairs is used to describe the size of a DNA molecule.

Chromosome: The self-replicating genetic structure of cells containing the cellular DNA that bears in its nucleotide sequences the linear array of genes. In prokaryotes, chromosomal DNA is circular, and the entire genome is carried on one chromosome. Plant genomes consist of a number of chromosomes whose DNA is associated with different kinds of proteins.

Clone: An exact copy made of biological material such as a DNA segment (a gene or other region), a whole cell, or a complete organism.

Cloning Vector: A piece of DNA, such as a plasmid, into which a DNA segment can be inserted, transferred into an organism, and replicated or reproduced.

DNA (deoxyribonucleic acid): The molecule that encodes genetic information. DNA is a double stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases: adenine (A), guanine (G), cytosine (C), and thymine (T). In nature, base pairs form only between A and T and between G and C; thus the base sequence of each single strand can be deduced from that of its partner.

EST: Expressed Sequence Tag: A unique, short DNA sequence derived from a cDNA library. ESTs are useful for localizing and orienting the mapping and sequence data reported from many different laboratories and serve as identifying landmarks on the developing physical map of a plant genome.

Expression Pattern: Gene expression is the process by which a gene's coded information is converted into the structures or molecules present and operating in the cell. Expression pattern refers to a set of genes expressed under a set conditions (e.g. genes expressed in plants grown under drought condition).

Functional Genomics: Studies of the relationship between the structure and organization of the genome and the function of the genome as it directs growth, development, physiological activities, and other life processes of the organism.

GenBank: A public database where DNA sequences are deposited and made public. It is operated and supported by the National Library of Medicine, part of the National Institutes of Health, and is part of an international consortium of a gene sequence database.

Gene: The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule).

Genetic Map: A map of the relative positions of genetic loci on a chromosome, determined on the basis of how often the loci are inherited together.

Genetics: The study of the patterns of inheritance of specific traits.

Genome: All the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs.

Genome Project: Research and technology development effort aimed at mapping and sequencing some or all of the genome of human beings and other organisms.

High Throughput Biology: An experimental approach that generates massive amounts of raw data at the production scale using highly automated technologies such as genome sequencing technology or microarray technology, and processes the data by a batch method using computational and other information management tools.

Human Genome Project: The national effort, led by DOE and NIH, was started in the late 1980's. It includes several projects to (1) determine the sequence of human DNA, (2) develop new computational methods for analyzing genetic map and DNA sequence data, and (3) develop new techniques and instruments for detecting and analyzing DNA. While the ultimate objective is to understand the structure, organization and function of the human genome, the Human Genome Project supports studies on several model microbial and animal genomes. No plant genomes are targeted by the Human Genome Project.

Informatics: The study of the application of computer and statistical techniques to the management of information. In genome projects, informatics includes the development of methods to search databases quickly, to analyze DNA sequence information, and to predict protein sequence and structure from DNA sequence data.

Library: An unordered collection of clones (i.e., cloned DNA from a particular organism), whose relationship to each other can be established by physical mapping.

Microarray Technology: New approach to the study of how large numbers of genes interact with each other. This technology provides a quantitative assessment of how a cell's regulatory networks control extensive gene sets simultaneously. The method uses a robot to precisely apply tiny droplets containing functional DNA to glass slides. Researchers then attach fluorescent labels to DNA from the cell they are studying. The slides are put into a scanning microscope that can measure the brightness of each fluorescent dot; brightness reveals how much of a specific DNA fragment is present, an indicator of how active it is.

Physical Map: A map of the physical locations of identifiable landmarks on DNA (e.g., restriction enzyme cutting sites, genes); distance is measured in base pairs. The highest resolution map would be the complete nucleotide sequence of the chromosomes.

Reverse Genetics: An experimental approach that begins with information about the primary DNA or protein sequence and uses this knowledge to generate targeted mutations (heritable changes) or altered expression levels. Observation of the resulting effects on the organism (a physical attribute such as the color of the flower or the shape of the leaf) yields information about the physiological function of the gene or protein. This is the reverse strategy to a classical genetics approach that proceeds from observing a defined genetic trait towards obtaining sequence information for a specific gene.

Sequencing: Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule or the order of amino acids in a protein.

Structural Genomics: Studies of the structure and organization of the genome including DNA sequencing and physical and genetic mapping.

Syntenic Map: Chromosomal maps showing the collinear relationships between the genomes of different organisms. Often these chromosomal regions from related organisms (such as various grass species) contain corresponding genetic information and similar gene order.

Technology Transfer: The process of converting scientific findings from research laboratories into useful products by the commercial sector.

Abstract

The National Plant Genome Initiative (NPGI) program was initiated in May, 1997 with the long-term objective to understand the structure and function of genes in plants important to agriculture, environmental management, energy, and health. This report highlights the progress made since the NPGI plan was implemented in FY1998, and describes future plans to achieve the six initiative goals. New projects to develop research resources for elucidating the sequence, structure, and organization of complex plant genomes have been initiated, including individual efforts and collaborations between universities, research institutes, private industry, and international research communities. Results from these efforts will provide a solid foundation and a springboard for plant biologists to extend traditional gene discovery and expression analyses to an integrated functional genomics approach. Through relating the function of individual genes to their impacts on plant production and productivity, new opportunities will exist for fundamental plant biology researchers as well for scientists with an interest in the transfer of genomics information into practical applications for plant-based industries. To maintain the momentum of the NPGI program, it is recommended that additional investments be made across all participating agencies at the level, for the next three-year period, of a total of \$500 million.

For further information, contact:

National Science and Technology Council Executive Secretariat at
(202) 456-6120 (voice) or (202) 456-6026 (fax).

Also available on the NSTC Home Page via link from the OSTP Home Page at:

http://www.whitehouse.gov/WH/EOP/OSTP/html/OSTP_Home.html



EXECUTIVE OFFICE OF THE PRESIDENT
Office of Science and Technology Policy
Washington, D.C. 20502